

THE HUMAN FACTORS OF CONDUCTING FIELD RESEARCH: LESSONS LEARNED FROM THE EVALUATION OF T-NASA

Anthony D. Andre
San Jose State University

Becky L. Hooley
Monterey Technologies, Inc.

This paper discusses the human factors of conducting field research, within the context of a recent field evaluation of a new aircraft taxi navigation and situation awareness (T-NASA) system developed at NASA Ames Research Center. Several critical issues that were confronted in the process of the field evaluation effort are presented, followed by an account of how these issues were resolved. From these experiences, guidelines concerning both the objectives and methods of field evaluations are provided to assist researchers in their future field evaluation efforts.

INTRODUCTION

Increasingly, human factors researchers are asked to address more applied issues and in many cases are involved in both design and validation phases (e.g., Koonce, Moore & Benton, 1995; Price & Dowell, 1997). Often, applied aviation HF projects start off with some form of field research aimed at identifying real-world issues and constraints (see Sarter & Woods, 1992; Andre, 1995; Kaempf, Klein, Thordsen & Wolf, 1996). The validation phase often involves field evaluations—semi-controlled tests of pilot-vehicle interfaces in actual field settings. Yet, the transfer of skills and issues from the laboratory research domain to that of the field study are often minimal.

THE SYSTEM UNDER EVALUATION

The system under evaluation is a new aircraft taxi navigation system, referred to as T-NASA (Taxi Navigation and Situation Awareness), aimed at improving the capacity and efficiency of aircraft ground operations in low-visibility conditions (Foyle, Andre, McCann, Wenzel, Begault, &

Battiste, 1996). T-NASA consists of two main components: 1) a panel-mounted electronic taxi map display which depicts the ownship, the airport surface and other traffic on a track-up, perspective moving map display; and 2) a heads-up display (HUD) which presents conformal, scene-linked symbology depicting the centerline and edges of the taxiways and runways (see Foyle, Andre, McCann, Wenzel, Begault, & Battiste, 1996). These components were installed in NASA's B757 research aircraft (see Figures 1 and 2).

Four B757 captains from four different airlines, each teamed with NASA co-pilots, participated in the field evaluation. All trials were conducted at night, under VFR visibility conditions, and in the context of normal airport traffic/ground control handling at Atlanta Hartsfield Airport (for more details, see Young and Jones, 1998).

Each pilot performed five trials with both the electronic map and the HUD, one trial with only the electronic map and one trial with only a conventional paper map. System and aircraft performance, map control inputs, workload ratings, situation awareness ratings, surveys, pilot debrief interviews and observational data were collected and analyzed.



Figure 1. NASA B757 with HUD and map.

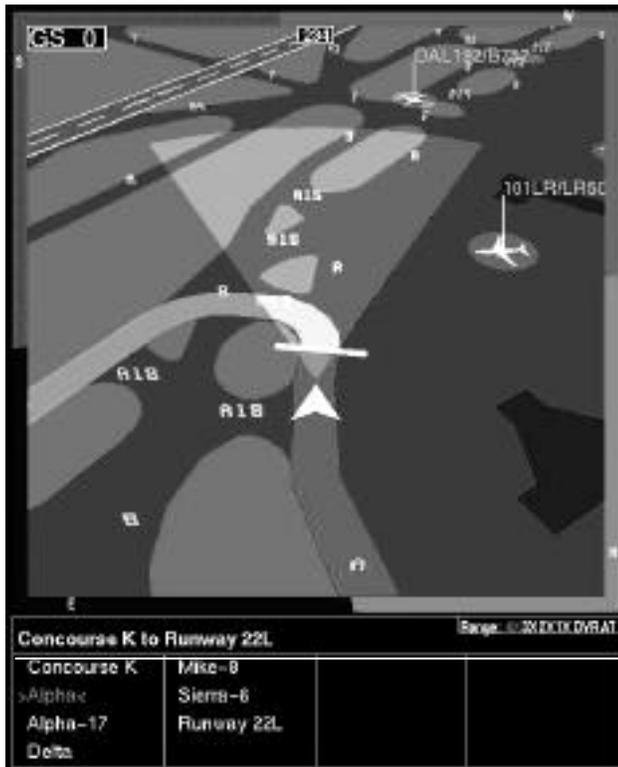
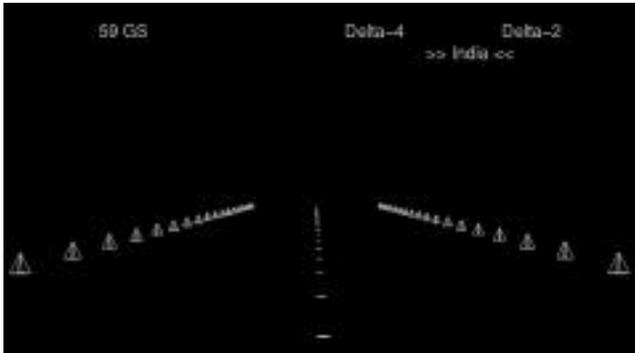


Figure 2. T-NASA HUD (top) and electronic map (bottom) displays.

THE NATURE OF FIELD RESEARCH

Much has been written about the many differences that exist between laboratory settings that vary in fidelity, for example, between various types of flight simulators (Koonce, 1979). It is often the case that as the fidelity of the laboratory evaluation increases, the experimental control decreases. Further, when the evaluation is conducted in the field, the notion of experimental control, as well as a host of other issues, might be altered dramatically.

Before discussing the issues involved in the conduct of field research, it's important to define the reasons that field research is conducted in the first place and to categorize the different forms of field research.

Why Field Research? Increasingly, as aviation HF researchers are asked to evaluate and design applied systems, the need to ground their work with field observation and evaluation increases (Hutchins, 1995). By applying a more ecological approach to our work, whereby the issues, users, dynamics, constraints and environmental factors are observed in real time, we can better apply human factors data and methods to the evaluation, design and validation process (e.g., see Kaempf, Klein, Thordsen & Wolf, 1996). The complexity and required flexibility of modern user interfaces means that we can no longer sit in our offices and make meaningful commentary, let alone design or training decisions, regarding the utility of advanced cockpit interfaces.

Types of Field Research. Typically, field research falls into 2 categories. Field ethnography describes the general activities that involve learning how pilots perform their job and the critical design or procedural issues that affect them. Field evaluation and validation describes the general activities involved in testing the utility of a new or revised cockpit interface design or procedure.

FIELD RESEARCH ISSUES

A number of issues had to be dealt with before, during and after the T-NASA field evaluation,

including: research objectives, research focus, experimental design, environmental factors, pilot instructions and behavior, pilot training, data collection, confounds and more. Only the most important of these issues are discussed below, accompanied by corresponding guidelines to aid future researchers in the conduct of field evaluations/validations.

Research Objectives

Decisions must be made up front as to the objectives of the field research, as all other decisions hinge on this. At a global level, one must decide if they are conducting an evaluation or a validation, as each requires different assumptions, approaches and measures. At a local level, the researcher must confront the differences between design research and usability testing.

In the lab, researchers are typically focused on proving the value of their designs, and on supporting their design decisions with established theory. They accomplish these objectives through controlled studies, often involving many factors.

In the T-NASA field study we first attempted to conduct a validation of the system, based on the stated taxi efficiency goals of the TAP program. However, we soon realized that both relevant real-world conditions would not be present and artificial conditions would be introduced; thus, we ended up with more of an evaluation than a validation.

Guidelines. 1) The term validation implies a higher level of fidelity and a low level of experimental control. If your objective is to validate the utility of a new display, then make sure that the fidelity of the operating context of your study is maximized. 2) The main purpose of field research should not be to prove the value of some design concept, but rather to determine if it can be successfully utilized in the actual context for which it is intended. 3) In addition, rather than using this rare opportunity as yet another way to demonstrate our design and human factors skills, the other main objective of field research should be to learn about potential usability and integration problems. You

will be surprised to learn how many “real-world” issues, never considered in the bowels of our human factors laboratories, have a large impact on the ultimate usability of a cockpit display or system.

Research Focus

We often evaluate our designs somewhat isolated from the physical and procedural implementation issues. Yet, at the point in which a field evaluation is undertaken, it is usually the case that we know a lot about the potential value of our design per se, but little about the hardware, software, training and procedural requirements and constraints.

In the T-NASA study, we were unable to assess the physical requirements, as the system was not evolved to a point where it could be implemented and integrated as a true avionics component. However, by including researchers who specialize in the study of crew roles and procedures, we were able to glean relevant training and procedural issues, constraints and requirements.

Guidelines. 1) While it is important to assess the value of the concept and features of the display in question, the field test should be constructed in a way that allows the assessment of the physical and procedural constraints and requirements. After all, the system is not designed for the lab, it's designed for the cockpit. 2) Researchers who specialize in the study of crew roles and procedures should be included in the field evaluation, and arrangements should be made so that they can observe the flight crews and record flight crew interactions via video and audio tape.

Experimental Design

All researchers have to struggle with the issue of how many factors to study and with how many levels. This decision is made more critical in the field for the reason that some factorial or nested designs might run the risk of not having enough data or equal amounts of data to analyze the factors in question.

The design of the T-NASA study at first contained too many factors and levels, and the order of trials did not guarantee that any main effects, let alone interactions, could be analyzed. While it may not conform to conventional counterbalance techniques, it's important to construct trials in an order that assumes that some will not be completed.

Guidelines. 1) Don't treat a field study like a lab study, since in the field, you really don't control anything. 2) Minimize the number of factors and levels manipulated. 3) Don't worry about experimental power. Rather, try to expose your design to as many real-world conditions as possible. Remember, you should not be trying to *prove* anything; you should instead be trying to *learn* about the real-world utility of your system.

Environmental Factors

We can't control the weather or other environmental factors, so these must be carefully considered in the design of the protocol. While day/night comparisons can easily be made, controlled comparisons of many other environmental factors will be difficult to plan.

In the T-NASA study we initially included different time and visibility conditions. However, these conditions did not avail themselves and we had to limit the remaining trials to nighttime only.

Guidelines. 1) If weather or visibility conditions are unpredictable, don't manipulate these factors.

Safety Regulations

As UFO enthusiasts have known for a long time, "we are not alone." So too with field studies, as the human factors researcher must work within the objectives and constraints of other members of the field study team. For example, all real-world systems have safety rules and regulations, especially aviation. So, when conducting field evaluations using actual aircraft and operating in actual surface or airspace, one must confront a variety of safety issues that don't exist in the laboratory.

A variety of safety issues impacted the T-NASA study, the most critical revolving around the display implementation. An external display had to be used for the electronic map so as not to interfere with the software driving the EHSI, and the location of the external map display could not obstruct any primary instruments. The result was a location biased to the Captain's side, as shown in Figure 1.

Guidelines. 1) Before you finalize your experimental plans, share them with all other teams involved so that you uncover any conflicts ahead of time. 2) Remember that safety always comes first.

Participant Training

We often try to maximize the number of experimental trials at the expense of practice or training time. However, when assessing the impact of a new system in the field, it is vital that the pilots are provided sufficient training to incorporate the new displays into their normal visual scan patterns.

In the T-NASA study the pilots received extensive part-task and field training before performing in the experimental trials.

Guidelines. 1) Provide ample training so that the pilots both understand how to use the displays, and incorporate them into their visual scan patterns.

Participant Behavioral Mode

When pilots participate in research studies it is natural that they want to convey all of their experiences and opinions; indeed, these are often the most valuable data. However, when pilots provide a running commentary, the very "real-world" conditions under study can be compromised.

In the T-NASA study, the pilots first assumed we wanted them to act as test pilots, providing a running commentary during the trials. But this was the opposite of our objective—to evaluate the T-NASA system under "normal" operating conditions. So we asked the pilots to hold their thoughts and comments until the debriefing session, which occurred in the cockpit immediately after the flight.

Guidelines. 1) In most cases, it is best that the pilots do not provide a running verbal commentary; rather they should attempt to ignore the goals of the researcher, and instead concentrate on their flight tasks and responsibilities. 2) Debrief the pilots immediately after the flight, preferably in their operating context (i.e., the cockpit).

Data Collection

In the field, we often do not have the luxury to collect quantitative data, via a computer, and even when the opportunity exists, the data can be confounded by other factors.

In the T-NASA study, quantitative data on taxi speeds and accuracy was confounded by other factors not under our control (e.g., other traffic, ground control instructions). We therefore put a large emphasis on post-trial ratings, comparative surveys, video analyses and debrief interviews to develop a more comprehensive picture of the utility of the system under real-world conditions. But don't assume that pilots always prefer what they perform best with (see Andre and Wickens, 1995).

Guidelines. 1) Don't rely on quantitative data collected by computers. 2) Incorporate a variety of measurement and observational techniques.

Hidden Confounds

As noted above, in field research the opportunity for hidden and unintended confounds imposed by the testing situation is great. Often they may only be discovered after testing is complete.

Exploratory analyses of the T-NASA study data revealed a hidden confound created by changing clearances. This created a situation in which trials in one condition (HUD trials) were shorter and had more turns than trials without the HUD. The result was that pilots were forced to taxi slower on HUD trials, not because of the HUD's utility but because there were more turns in these taxi routes. Without

discovering this confound, the data may have shown an artificial cost of taxiing with a HUD.

Guidelines. 1) Don't assume that the factors and conditions manipulated were unconfounded by other factors, out of your control. 2) During the study, note any circumstances that may confound your manipulations/comparisons. 3) After the study, expect and look for confounds in the data.

SUMMARY

In summary, the nature of aviation human factors research has shifted from a basic, laboratory mode to one in which more applied design problems are addressed and subsequently supported and evaluated in field settings. This leaves the typical researcher with a set of skills and experiences that might not transfer well from the lab to the field setting. Many researchers are now utilizing field ethnography methods to determine critical operational issues (e.g., Sarter and Woods, 1992; Kaempf, Klein, Thordsen & Wolf, 1996), and some are starting to evaluate and validate their proposed designs in field settings as well.

In the context of a recent field evaluation of a new taxi navigation system (T-NASA), we have discussed several issues that underlie the successful conduct of field research and proposed guidelines for future researchers who engage in field evaluations and validations. We are confident that such information will become increasingly needed by the aviation HF community, and we hope other researchers will be equally motivated to share their experiences in the field.

REFERENCES

- Andre, A.D. (1995). Information requirements for low-visibility taxi operations: What pilots say. *Proceedings of the 8th International Symposium on Aviation Psychology*, pp. 484-488, Columbus, OH.
- Andre, A.D. and Wickens, C.D. (1995). *When users want what's not best for them*. *Ergonomics*

- in Design (pp. 10-14). Human Factors & Ergonomics Society, Santa Monica, CA.
- Foyle, D.C., Andre, A.D., McCann, R.S., Wenzel, E., Begault, D. and Battiste, V. (1996). Taxiway Navigation and Situation Awareness (T-NASA) System: Problem, design philosophy and description of an integrated display suite for low-visibility airport surface operations. *SAE Transactions: Journal of Aerospace*, 105, 1411-1418.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, Massachusetts: MIT Press.
- Kaempf, G.L., Klein, G.A., Thordsen, M.L. & Wolf, S. (1996). Decision making in complex Naval command-and-control environments. *Human Factors*, 38(2), 220-231.
- Koonce, J.M. (1979). Predictive validity of flight simulators as a function of simulator motion. *Human Factors*, 21, 215-223.
- Koonce, J.M., Moore, S.L., & Benton, C.J. (1995). Initial validation of a basic flight instruction tutoring system (BFITS). *In Proceedings of the Eighth International Symposium on Aviation Psychology*, pp. 1037-1040. Columbus, OH: OSU.
- Price, J.M., & Dowell, W.R. (1997). A field evaluation of two split keyboards. *In Proceedings of the HFES 41st Annual Meeting*, pp. 410-414. Santa Monica, CA: HFES.
- Sarter, N.B. and Woods, D.D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *International Journal of Aviation Psychology*, 2(4), 303-321.
- Young, S., and Jones, D. (1998). Flight Testing of an Airport Surface Movement Guidance, Navigation, and Control System. *Proceedings of the Institute of Navigation's National Technical Meeting*.